Autocorrelation

Dr. Zhi Wang
Department of Earth and Environmental Sciences
California State University, Fresno

Descriptive

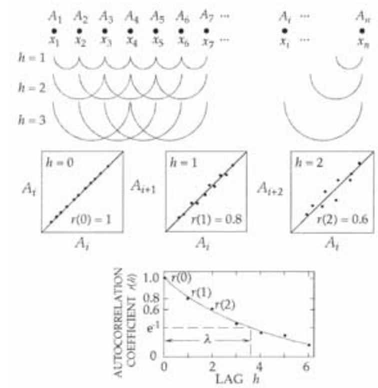
- Not to determine the relation between two different variables, but the relation between only one variable observed at different locations or times.
- For measurements taken very close together, their values would be alike while those separated by a great distance (time) could differ markedly.

Relevant questions

- Are the measurements spatially or temporally correlated?
- Are they spatially or temporally independent of each other?
- How large area (what time period) does a measurement represent? How to improve measurement?
- How large (often) should a sample be taken?
- Is there a relation between sample size and autocorrelation length?

Spatial description

- The distance between neighbors h is called the lag (?).
- The distance over which a significant correlation exists is called the autocorrelation length λ.



Calculating an autocorrelation function for soil property A along a transect.

Autocorrelation Coefficient

covariance:

$$\mathrm{cov}_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left[x_i-\bar{x}\right]\left[y_i-\bar{y}\right], \quad r_{xy} = \frac{\mathrm{cov}_{xy}}{s_x s_y}$$

auto covariance

$$\mathrm{cov}\left[A_i(x), A_i(x+h)\right] = \frac{1}{N}\sum_{i=1}^{N-h}\left[A_i(x_i)-\bar{A}\right]\left[A_i(x_i+h)-\bar{A}\right]$$

autocorrelation coefficient

$$r(h) = \frac{\mathrm{cov}\left[A_i(x), A_i(x+h)\right]}{\sqrt{\mathrm{var}\left[A_i(x)\right]}\sqrt{\mathrm{var}\left[A_i(x+h)\right]}}$$

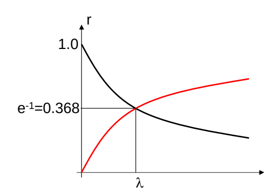
Correlation Length λ

autocorrelation coefficient and λ

$$r(h) = r_0 \exp(-h/\lambda), \quad r_0 = r(h=0) = 1.0$$

λ is obtained from

$$\sum_{i=1}^{n-1}\left[r_i(h) - \exp(h/\lambda)\right] = 0$$



Significant Confidence Interval

Is the autocorrelation coefficient r(h) significantly different from zero?

- Yes, if the autocorrelation length λ>0.
- Yes if the following Z-score is less than the critical $Z_{1-\alpha/2}$.

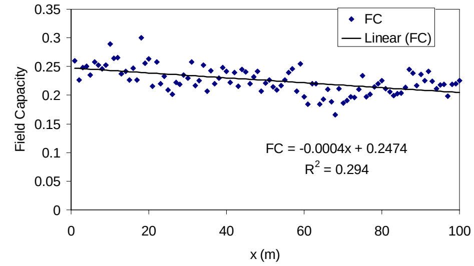
$$Z = r(h)\sqrt{(n-h+3)}$$

n is the number of observations.

Excel function: $Z_{1-\alpha/2}$=NORMSINV(1-α/2)

$if\ 1-\alpha = 95\%,\ Z_{\alpha/2} = -1.96,\ Z_{1-\alpha/2} = 1.96$

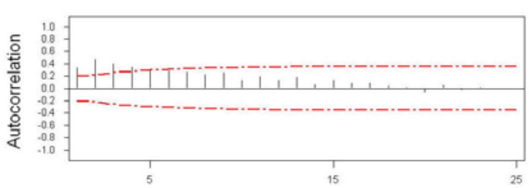
Autocorrelation Values



Autocorrelation Values

Autocorrelation is a <u>systematic</u> pattern in the errors that can be either attracting (**positive**) or repelling (**negative**) autocorrelation.

Autocorrelation Function for FC



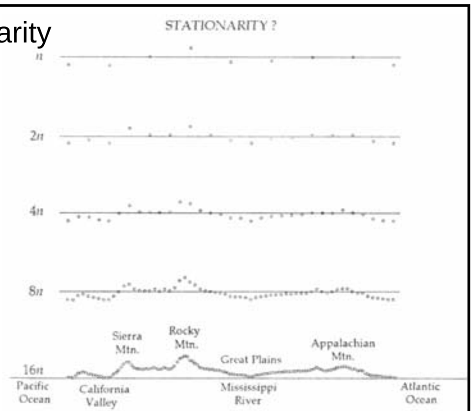
In General

- Physical, chemical and biological reactions occurring rapidly manifest short correlation lengths than do those occurring more slowly.
- Hence, sampling frequencies should be higher for rapidly changing variables and lower for slowly varying ones.

Stationarity and Trend

- The usual Assumption for autocorrelation analysis is that the mean value (of the population) exists and stable for each set of sample observations.
- Actually, the mean of a population is a function of location, sampling distance and volume, as well as the size of the domain being sampled.
- Stationarity (static true condition) can be reached by adjusting the above factors.
- Since the natural systems maybe in a transient state, stationarity reflects a periodic trend.

Stationarity



Stationarity versus number of observations over the same domain

Deterministic Trend

- Deterministic and random variations of the landscape attributes are often observed simultaneously.
- The deterministic trends need to be removed before autocorrelation analysis.
- Two Detrending methods:
  - Remove linear autocorrelations by subtracting from r(h).
  - Taking the differences of the neighboring measurement for autocorrelation.

Computer Exercises

- Use MiniTab and the provided EXCEL dataset to find autocorrelograms of soil attributes.
- Determine the correlation length λ and the 95% confidence intervals.